

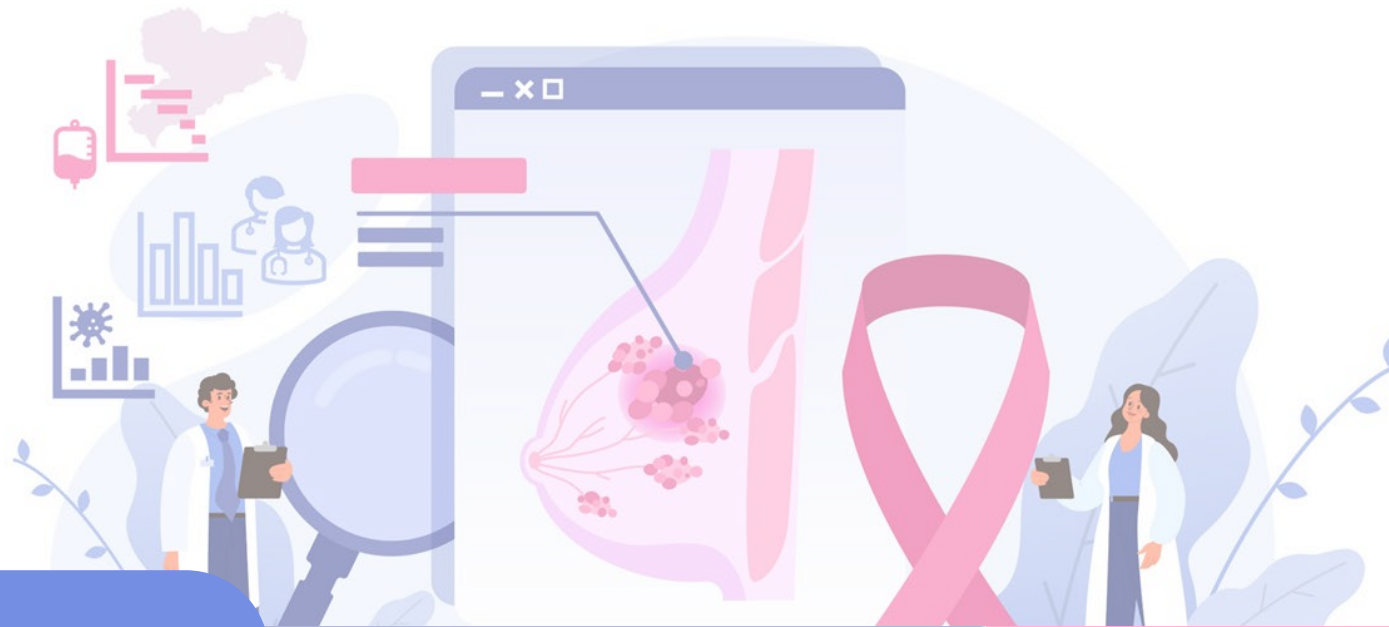
Gefördert durch

STAATSMINISTERIUM FÜR  
SOZIALES, GESUNDHEIT UND  
GESELLSCHAFTLICHEN ZUSAMMENHALT



# Behandlungsmigration verstehen

CHEMNITZ | 22. APRIL 2026 | DR. PHILIPP SCHULZ | DATA SCIENTIST



VERSORGUNGSFORSCHUNG

BEHANDLUNG DES  
MAMMAKARZINOMS

AKTUELLE INITIATIVEN  
IN DER VERSORGUNG

## Fragestellung:

Welche räumlichen Muster der operativen Versorgung sind für Fälle des Mamma-Karzinoms in der Gesundheitsregion Südwestsachsen erkennbar?

# Inhalte

---

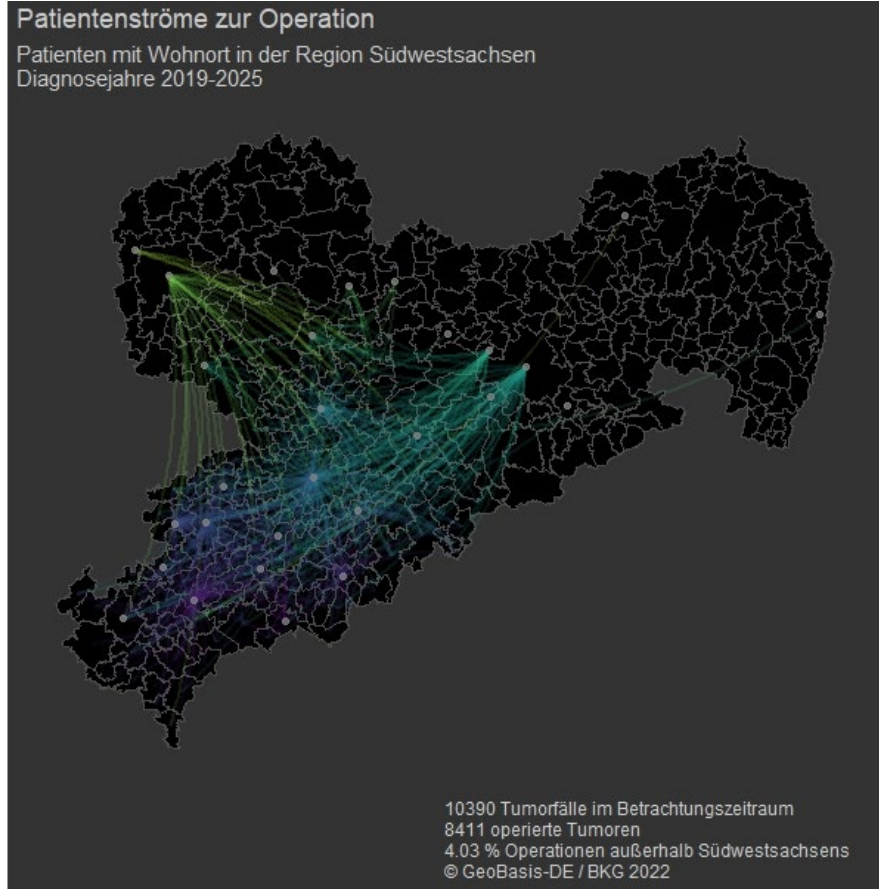
1. Selektionskriterien
2. Deskriptive Kennzahlen
3. Bivariate Zusammenhänge
4. Machine-Learning Klassifikationsmodell
5. Zusammenfassung und praktische Implikationen

# 1. Selektionskriterien

---

- Analyseebene Tumor
- männliche und weibliche Patienten mit Mamma-Karzinom (C50, D05)
- Diagnosejahre 2019 – 2025
- Wohnort zum Zeitpunkt der Diagnose in den Landkreisen Chemnitz, Erzgebirgskreis, Mittelsachsen, Vogtlandkreis, Zwickau
- Einschluss [tumorrelevanter Therapien](#)
- Betrachtung des ersten operativen Eingriffs

# 1. Deskriptive Kennzahlen

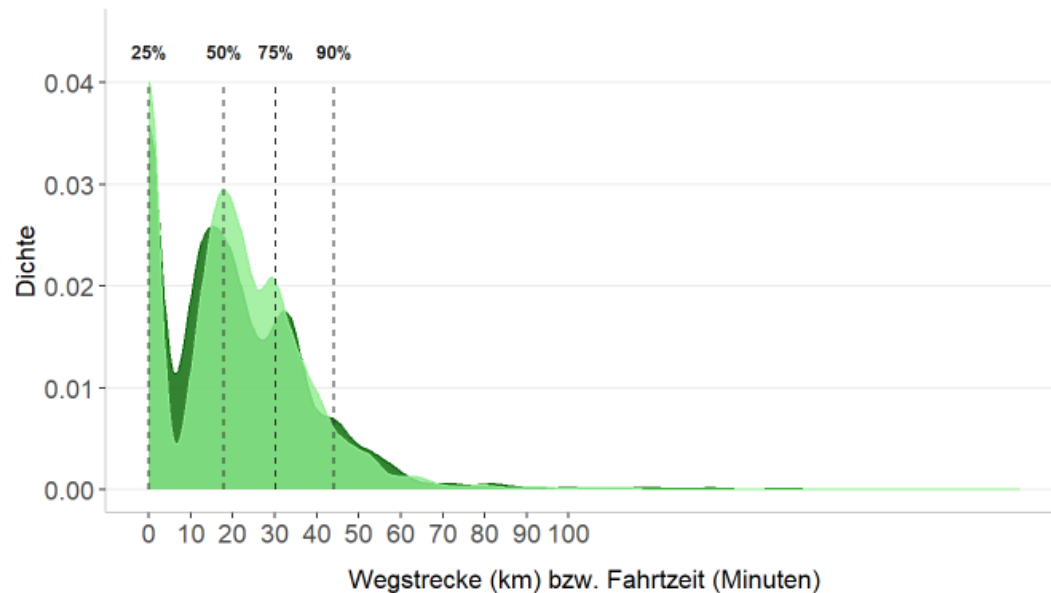


	min	max	median
Distanz kürzeste Strecke (km)	0	192,3	17,0
Distanz schnellste Strecke (km)	0	197,7	17,8
Fahrtzeit kürzeste Strecke (Minuten)	0	152,8	18,3
Fahrtzeit schnellste Strecke (Minuten)	0	154,1	18,3

# 1. Deskriptive Kennzahlen

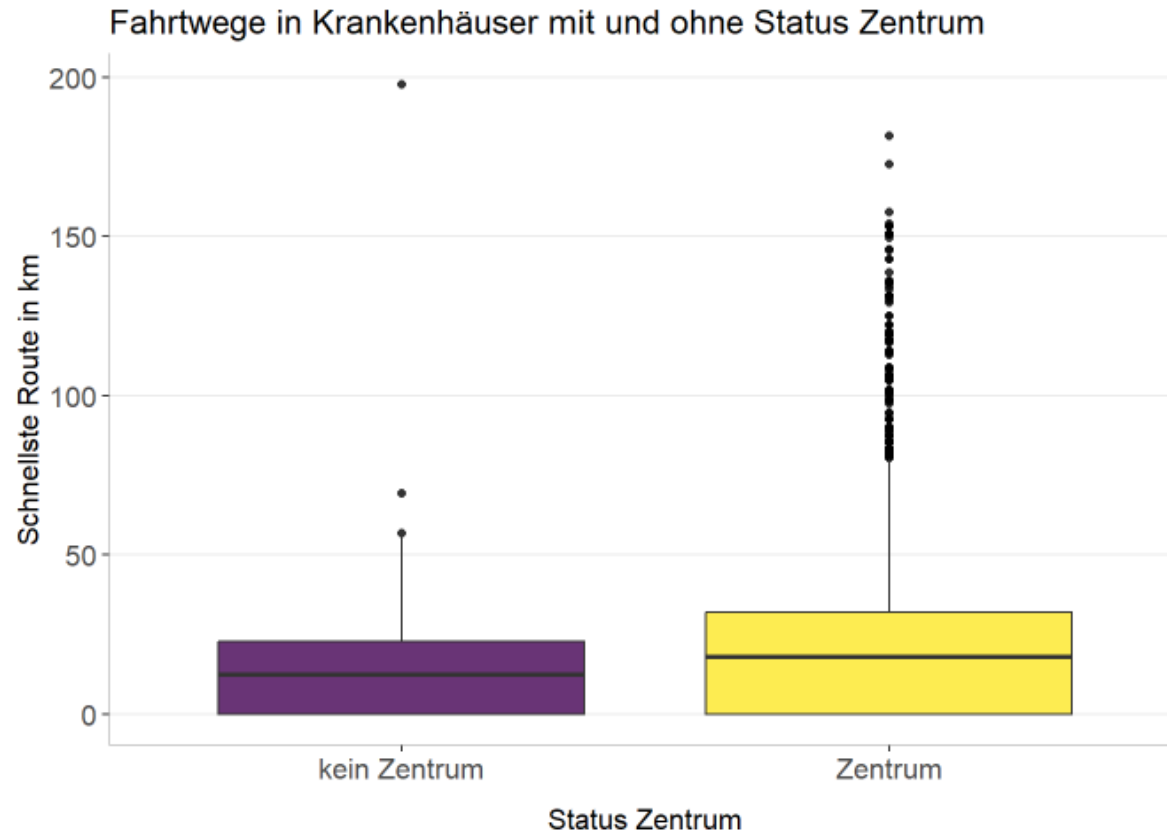
Distanz und Fahrtzeit zum Operationsort  
schnellste Route

■ Distanz (km) ■ Fahrtzeit (min)



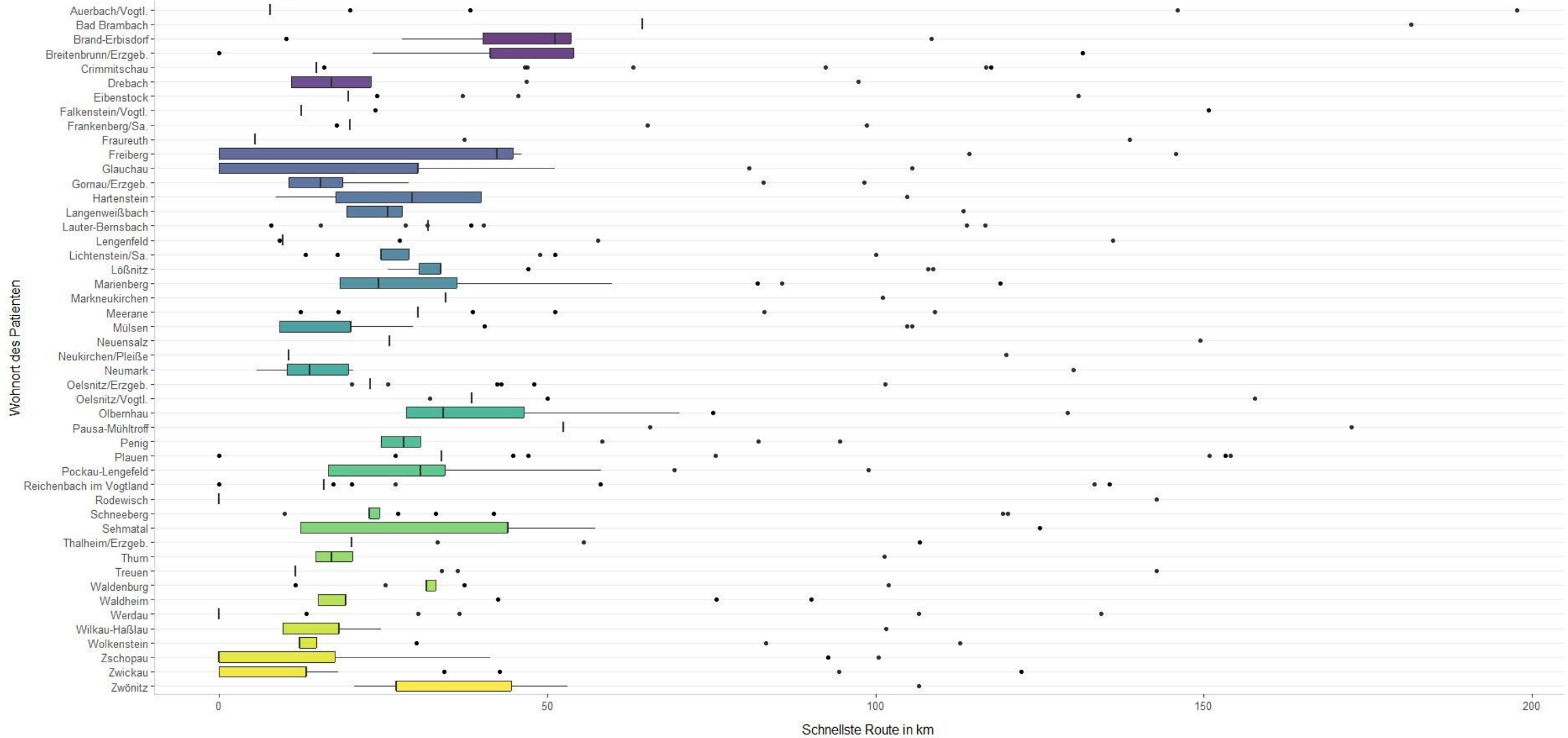
- 25 % der Tumorfälle mit OP am Wohnort
- 50 % in einer Distanz von weniger als 20 km
- 75 % in einer Distanz von 30 km
- 90 % in einer Distanz von weniger als 50 km

### 3. Bivariate Zusammenhänge

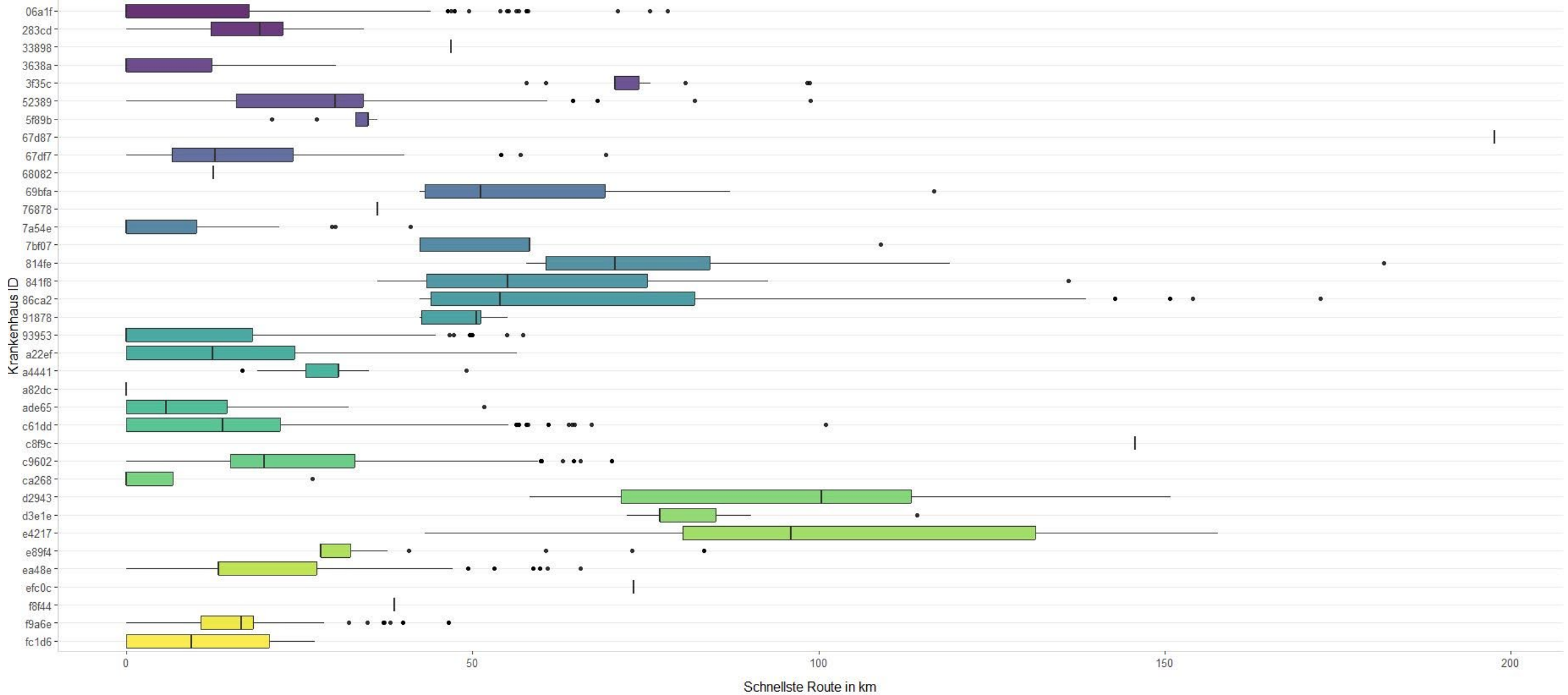


### Fahrtwege nach Wohnort des Patienten

Wohnorte mit Fahrtwegen über 90 km



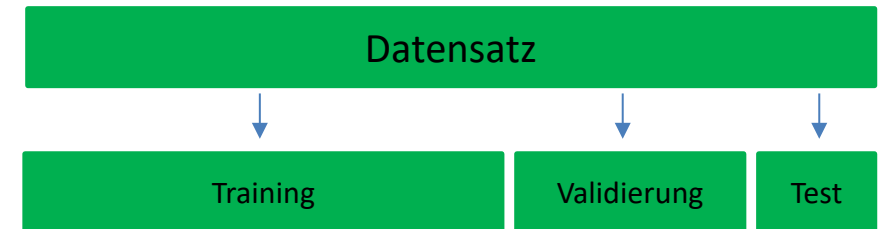
### Fahrtwege je Krankenhaus-ID



## 4. Machine-Learning Klassifikationsmodell

---

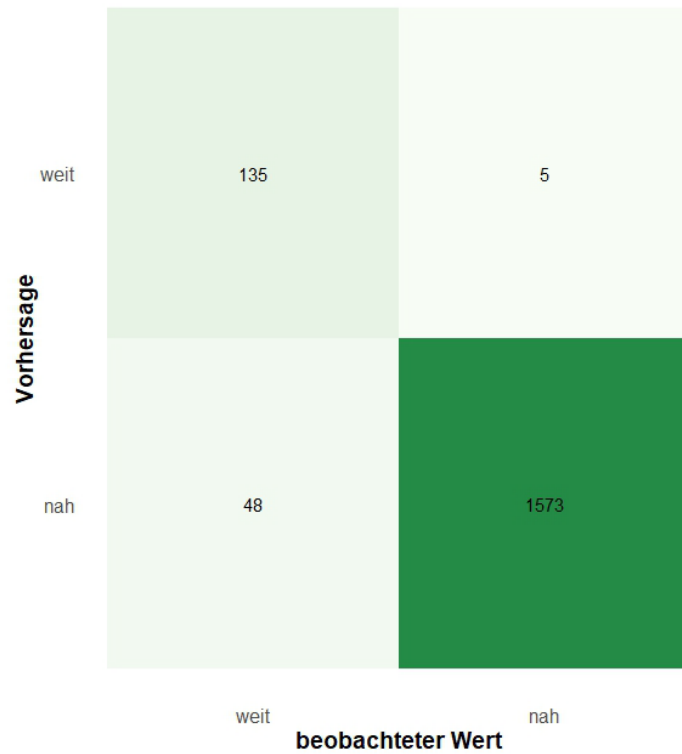
- **Motivation und Modellauswahl:**
- deskriptive und bivariate Verfahren bringen keine neuen Erkenntnisse
- kombinierten Effekt von Einflussfaktoren möglichst unkompliziert abbilden
- LightGBM- Algorithmus
  
- **Vorgehen:**
- Definition Label für Klassifikationsaufgabe
- Definition von Features
- Train-Test-Split
- Grid-Search zur Optimierung der Modellparameter
- Modellevaluation



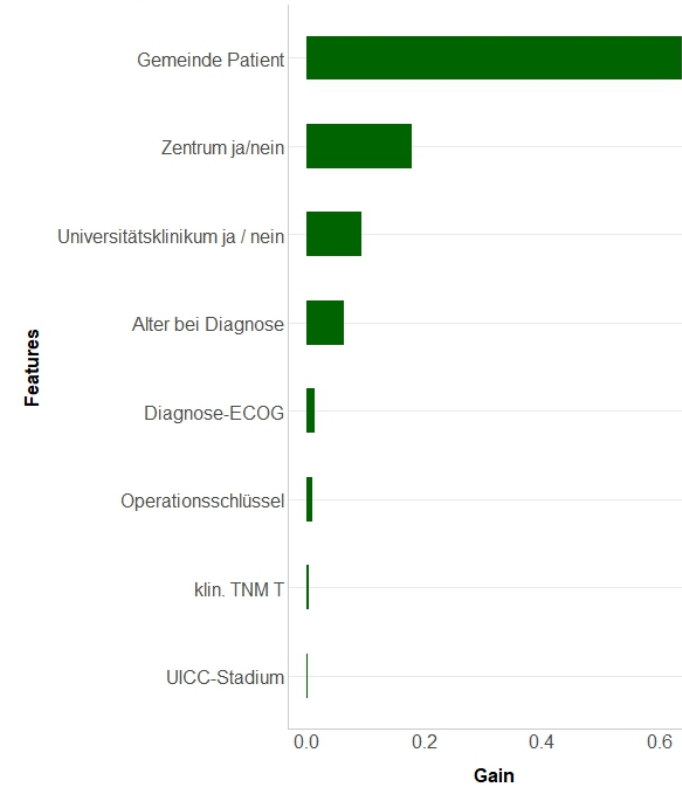
# 4. Machine-Learning Klassifikationsmodell

**Konfusionsmatrix**

Top 10 % der Fahrtstrecke zur OP = weit



**Feature Importance**



## 5. Zusammenfassung und praktische Implikationen

---

- **Versorgung ist flächendeckend gesichert:**
- 90% der Patienten fahren weniger als 41 km zur Operation
- Weitfahrer kommen überwiegend aus dem ländlichen Raum und lassen sich in Zentren/Universitätskliniken behandeln
  
- **Keine Hinweise auf Spezialisierung von Kliniken:**
- verwendete klinische Features tragen nicht nennenswert zur Fehlerreduktion des Modells bei
- verwendete Features zu Operationen tragen nicht nennenswert zur Fehlerreduktion des Modells bei



Vielen Dank für Ihre Aufmerksamkeit

### **Kontakt**

Dr. Philipp Schulz

Data Scientist

[p.schulz@krebsregister-sachsen.de](mailto:p.schulz@krebsregister-sachsen.de)

# Backup - Limitationen und Ausblick

---

- **Vollständigkeit der Fälle:**
  - insb. Behandlung in anderen Bundesländern
  - Behandlung von Patienten aus anderen Regionen
- **Erweiterungen im Feature-Set:**
  - Charakteristika von Zuweisern, prä-operative Therapien, persönliche Merkmale Patient, Leistungsportfolio von Kliniken, ...
- **Betrachtung ergänzender Fragestellungen:**
  - Definition anderer Labels (Behandlung in Zentren, Behandlung in Universitätskliniken, ...)
  - Betrachtung von Unterschieden in den Outcomes
  - Betrachtung räumlicher Muster post-operativer Therapien

# Backup – Grundidee LightGBM

---

- **LightGBM-Algorithmus funktioniert wie ein Team aus vielen Experten:**
- erster Experte schätzt fehlerhaften (!) Entscheidungsbaum d.h. versucht Daten optimal zu splitten, um Klassifikationsaufgabe zu lösen
- zweiter Experte fokussiert sich nur noch auf die Fehler der Schätzungen des ersten Experten, um diese zu korrigieren
- ...
- Fokussierung auf wesentliche Features, unwesentliche Datenpunkte werden ignoriert
- hohe Effizienz bei Berechnung
- **Kombination aller Schätzungen zu einem extrem genauen Ergebnis**